11. Single-Photon Generation in a Single-Photon Turnstile Device

11.1 Introduction

The model and discussion presented in Chap. 10 demonstrate the possibility of observing the Coulomb blockade effect in mesoscopic p–n junctions and further imply the possible realization of a "single-photon turnstile device" that generates single photons with a well-defined time interval. Since the arrival time of each photon can be predicted, such single photon states are called "heralded single photons" [104, 269]. This chapter describes the experimental progress towards realization of such single-photon states using a mesoscopic semiconductor p–n junction. The first half of the chapter discusses the fabrication of a mesoscopic p–n junction in a GaAs material system. The second half describes the electrical and optical characteristics of these devices measured at low temperatures. A Coulomb staircase was observed in the electrical measurement, which clearly demonstrates the Coulomb blockade effect present in these devices. Device operation in the turnstile mode indicates that the time interval between single photons is regulated to better than the Poissonian limit. This implies that nonclassical single photons can be generated from a mesoscopic p–n junction utilizing the Coulomb blockade effect.

11.2 Device Fabrication

Fabrication of sub-micron size p–n junctions in the GaAs system that could be used as a light emitter is a challenging proposition for several reasons. The general strategy would be to first grow a vertical p–n junction structure, and then pattern small posts using microfabrication technologies. The pattern can be generated using standard electron-beam lithography technology, but directional (anisotropic) etching with low damage and high spatial resolution is the first challenge. A second challenge is to make an electrical contact to the top of these individual, sub-micron posts and make the contact transparent to optical photons. The third and most difficult challenge is to ensure that the internal quantum efficiency is high enough for an electron–hole pair to produce a single photon in such a small device. This is difficult because it is

J. Kim et al., *Nonclassical Light from Semiconductor Lasers and LEDS*
© Springer-Verlag Berlin Heidelberg 2001

Table 11.1. Detailed structure of the double-barrier p–i–n junction grown for the single-photon turnstile device.

Type	Material	Role of Layer	Al Comp.	Doping density	Thickness(Å)
p+	GaAs	Cap layer	0	5×10^{19} cm^{-3} (max. doping)	100
p+	$Al_xGa_{1-x}As$	Contact layer	$x = 0.16$	5×10^{19} cm^{-3} (max. doping)	4000
p	$Al_xGa_{1-x}As$	p-lead	$x = 0.16$	5×10^{18} cm^{-3}	2000
i	$Al_yGa_{1-y}As$	p-side QW	$y = 0.12$	0	75
i	$Al_zGa_{1-z}As$	Barrier	$z = 0.18$	0	200
i	GaAs	Central QW	0	0	100
i	$Al_zGa_{1-z}As$	Barrier	$z = 0.18$	0	200
i	$Al_yGa_{1-y}As$	n-side QW	$y = 0.12$	0	90
n	$Al_xGa_{1-x}As$	n-lead	$x = 0.16$	2×10^{17} cm^{-3}	10000
n+	GaAs	Buffer layer	0	5×10^{18} cm^{-3}	5000
n+	GaAs	Substrate	0	5×10^{18} cm^{-3}	

known that the nonradiative recombination of an electron–hole pair through surface states in GaAs becomes a dominant mechanism in the radiative recombination as the device's size shrinks [155]. This section describes how a mesoscopic p–n junction was fabricated, and how these challenges were met.

11.2.1 Wafer Design and Growth

The resonant tunneling rates of electrons and holes for the structure shown in Fig. 10.1a can be calculated as discussed in Chap. 10. The thicknesses of the three quantum wells (QWs) and the two tunnel barriers need to be determined carefully, so that the electron and hole resonant tunneling conditions are well-separated in bias voltage and their respective tunneling rates are reasonably large. Table 11.1 shows the layers of the double-barrier p–i–n junction grown for the experiment. The wafer was grown using the molecular beam epitaxy (MBE) technique by our colleagues Dr. Kan and Mr. Watanabe, at Hamamatsu Photonics, Japan. We employed a digital alloying technique to form $Al_xGa_{1-x}As$ alloys for the tunnel barriers and alloy QWs for the p-side and n-side. This is where layers of AlAs and GaAs are alternately grown to form AlGaAs layers where the relative thickness of the layers determines the Al composition [219]. Such a digital alloying technique is known to provide better barriers than a random alloy, since random scattering in the alloy is replaced with more regulated interface scattering. It is known that, provided that the period of digital alloying is small enough, the photoluminescence linewidth of a QW sandwiched between digitally alloyed barriers features sharper linewidths.

For the structure above, one needs to estimate the bias voltage values where the electron and hole tunneling conditions are satisfied. This could be

done by solving the Poisson equation, given the doping concentration of each layer and the overall applied bias voltage. To be strict, one needs to find a self-consistent solution using both the Poisson equation and the Schrödinger equation, since the Poisson equation solves the voltage profile given the charge distribution, while the Schrödinger equation solves the electron wavefunction (which gives the charge distribution) given the voltage profile. Such a self-consistent solution is usually found using the mean field approximation (also known as the Hartree approximation). Rather than trying to solve these equations numerically, a very simple model was used here to estimate the resonant tunneling conditions. In the current structure, the carriers are supplied to the intrinsic region only by tunneling, and the number of carriers in this part of the junction is very small. The effect of electron tunneling into the central QW on the junction voltage profile is modeled by the charging energy. When an extra electron is in the central QW, the voltage drop across the tunnel barriers is modified as discussed in Chap. 10. Use of such a simple model is justified if electrons (or holes) are confined to the central QW in the intrinsic region and the number of charges that can be stored in the central QW is relatively small.

From the solution of the Poisson equation in the absence of any charge in the intrinsic region, one can calculate the bias voltage at which the first electron is allowed to resonantly tunnel into the central QW. The calculation is done assuming low temperature (which determines the bandgap of GaAs, and thus the built-in potential of the junction, and makes the Fermi function a step function) and yields the value $V_{\text{ext}} = 1.526$ V. Similarly, a first hole resonant tunneling is allowed at $V_{\text{ext}} = 1.589$ V, if no electrons are in the central QW. The difference between these resonant tunneling voltages is $\sim$63 meV.

11.2.2 Ohmic Contact Formation

The first step in the fabrication process is to form an n-type contact on the substrate side of the wafer. the n-type contact is formed by evaporating an appropriate alloy of Ni, Au and Ge [251]. A metal–semiconductor junction always forms a Schottky barrier in GaAs, since the Fermi level is pinned near the middle of the gap in GaAs. The carrier transport through a Schottky barrier is not Ohmic, i.e., the current–voltage characteristics are not linear, since the carriers need to be thermally injected above the Schottky barrier. The width of the Schottky barrier is determined by the doping density, N, of either donors or acceptors and is $\sim 1/\sqrt{N}$. The only way to make a reasonable Ohmic contact is to increase the doping density in the semiconductor near the contact and make the Schottky barrier thinner. When the barrier becomes thin enough, the electron (or hole) transport will be dominated by charge tunneling through the Schottky barrier, and the current–voltage characteristics become linear. For p-type contacts, one can usually dope p-type layers high enough ($\sim 10^{19}$ cm^{-3}) in the growth procedure so that this is not

an issue. For n-type layers, doping levels are limited to $\sim 10^{18}$ cm^{-3}, and good Ohmic contacts cannot be formed. One way to achieve high n-doping is to evaporate Au and Ge on GaAs and heat the sample up to ∼450°C. Ga atoms then diffuse into the Au, creating empty Ga sites, which become available for the Ge atoms to fill. Through such substitution, a very high n-type doping density can be achieved near the surface. Ni helps Ge atoms to diffuse into the GaAs crystal better, as well as playing the role of a wetting agent that prevents "balling up" of Au/Ge metal contacts. The temperature and duration of the anneal is a critical factor in determining the contact resistance. The anneal was done at 450 °C for 30 s in a rapid thermal annealer (RTA).

11.2.3 Device Definition: Electron-Beam Lithography

The electron-beam lithography technique was used to define a small junction area. Electron-beam lithography is a technique where a resist sensitive to electron-beam exposure is put down on the sample, and patterns are drawn by scanning a focused electron beam on the resist. Since a high-energy (30–40 keV) electron beam can have a significantly shorter wavelength (∼0.007 nm) compared to ultraviolet photons used in conventional optical lithography (∼250 nm), the diffraction-limited minimum feature size that can be defined with electron-beam lithography is much smaller than with optical lithography. The resolution of electron-beam lithography is practically limited by resist exposure due to secondary electrons that are emitted when a high-energy electron bombards the sample. There are several types of electron-beam resists available, both positive and negative. Polymethylmethacrylate (PMMA), a well-known positive resist with very high resolution, was used. PMMA is a large polymer molecule, and some bonds can be broken with electron-beam exposure. When the bonds are broken, the resulting smaller molecules can be dissolved by certain solvents (called developers). There are PMMAs with different molecular weights that have different development rates. Two types of PMMA were used, with molecular weights of 495 K and 960 K. For the developer, a 3:1 mixture of isopropanol and methylisobutylketone (MIBK) is commonly used. For electron-beam writing, a scanning electron microscope (Leica StereoScan 440) converted into an electron- beam writing system using external pattern generation and a scan control system (Nabity Systems) was used. The smallest linewidth we could achieve with this system was about 40 nm. Circular devices with radii between 200 and 1000 nm were defined using this process.

11.2.4 Metal Evaporation and Liftoff

The electron-beam lithography is performed in a way that is optimized for a metal liftoff technique. Figure 11.1 shows the liftoff procedure using electron-beam lithography. First, one spins on a layer of resist and patterns the resist

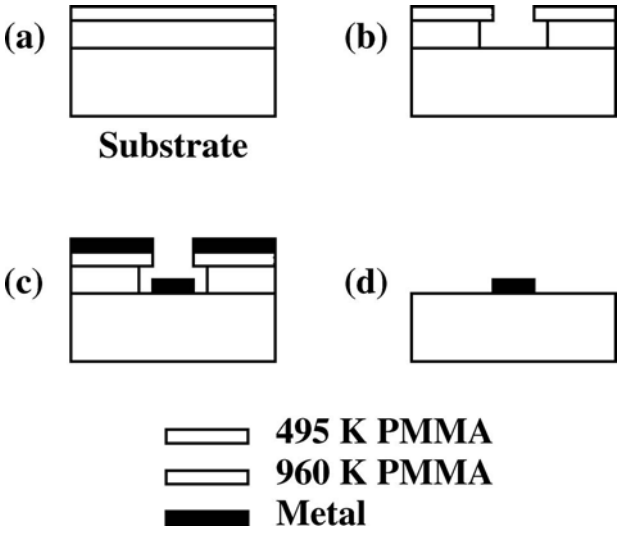

Fig. 11.1 a-d. Metal liftoff process using bilayer PMMA. **(a)** Bilayer PMMA film on a GaAs sample, **(b)** After electron-beam lithography, desired parts of the PMMA film have been removed. The use of bilayer PMMA creates an undercut profile in the resist, **(c)** Evaporation of metal. Line-of-sight deposition leaves patterns of metal on the wafer it are disconnected from the film on top of the resist, **(d)** After soaking in solvent, the resist and the metal film on top of that is removed and only the desired pattern of metal remains on the sample

to open up holes where one wants the metal to be put down. Then, a thin layer of metal is evaporated over the whole sample. Evaporation is a way to deposit thin films in a nonconformal way, i.e., metal atoms move in a straight path until they hit a surface where they stick. Therefore, the vertical side walls are not covered with metal. Then, the sample is soaked in a solvent that dissolves the resist. The metal sitting on top of the resist is removed from the sample, while the portion directly deposited on top of the sample remains. This way, one can leave a desired pattern of metal on the sample surface.

In order to enhance the liftoff process, it is very important that the metal film on top of the resist is not connected to the metal on the sample surface. Therefore, it is desirable to have a resist profile that features an undercut (overhang) structure [248]. In electron beam lithography, such a resist profile is naturally formed due to resist exposure by secondary electrons (called "proximity effect") [249]. Such undercut features can easily be enhanced further using a bilayer of PMMA. In this process, PMMA with a smaller molecular weight (e.g., 495 K) is put down on the bottom and PMMA with larger molecular weight (e.g., 960 K) is put down on top. It should be noted that a typical solvent for PMMA is chlorobenzene; however, this is such a good

solvent that an attempt to put down the second layer will completely dissolve the first layer. Therefore, 960 K PMMA in MIBK is used for the second layer. After exposure, the 495 K PMMA develops faster than the 960 K PMMA, and an undercut resist profile can easily be achieved. In the fabrication process discussed here, 100 Å of Cr was evaporated, followed by 300 Å of Au. This metal serves as a p-type contact, since the GaAs layer near the surface of the wafer is very highly doped.

11.2.5 Device Isolation: ECR-RIE

The metal circles left on the sample define the junctions. The next step is to isolate the devices from one another. This is done by a self-aligned etching process using a plasma. It is called "self-aligned" because no independent step is used to define the top contact: in this process, the Cr/Au metal circles serve both as the mask for etching and as the p-type ohmic contact. Since the p–i–n junction is located about 6000 Å below the surface (Table 11.1), the etch depth had to be at least 7000 Å. For the smallest devices defined, this would mean that we need etching techniques with an aspect ratio better than 5:1. Wet chemical etching cannot provide such high aspect-ratio etching; one has to use dry etching. The etching was done by reactive ion etching (RIE) in a high density of Cl_2/BCl_3 plasma generated by an electron cyclotron resonance (ECR) source. RIE is a dry-etching technique in which an etching ion (in this case Cl) chemically reacts with the GaAs crystal and forms a volatile species that can be pumped away in a vacuum chamber.

Conventional RIE utilizes the plasma sheath that develops between two parallel plates (or between a chuck that holds the sample and the remainder of the chamber) when a radio-frequency (RF) bias (at 13.56 MHz) is applied between the two plates [250]. In such RIE, the density of the plasma is determined by the applied RF power. At the same time, a DC voltage develops between the plasma sheath and the sample chuck (which is always used as one of the parallel plates) that accelerates the reactive ions towards the sample, and directional (anisotropic) etching can be performed. However, this DC voltage is also determined by the RF power and is on the order of 300 V. Such a high acceleration voltage induces serious damage in the sample due to ion bombardment, but is unavoidable since a certain RF power is required to generate any plasma at all.

ECR is one of many popular technologies to generate a very-high-density plasma. When a microwave source with a frequency of 2.45 GHz is applied to a gas in a magnetic field gradient, the cyclotron motion of the electrons (with angular frequency given by $\omega_c = eB/m$, where B is the magnetic field and m is the electron mass) becomes resonant with this microwave frequency when the magnetic field is 875 gauss. This means that at the plane where the magnetic field is 875 gauss (called the "resonant zone") the electrons absorb the microwave power resonantly and generate a very-high-density plasma. The pressure should be low enough ($1 \sim 20$ mtorr) so that the collision

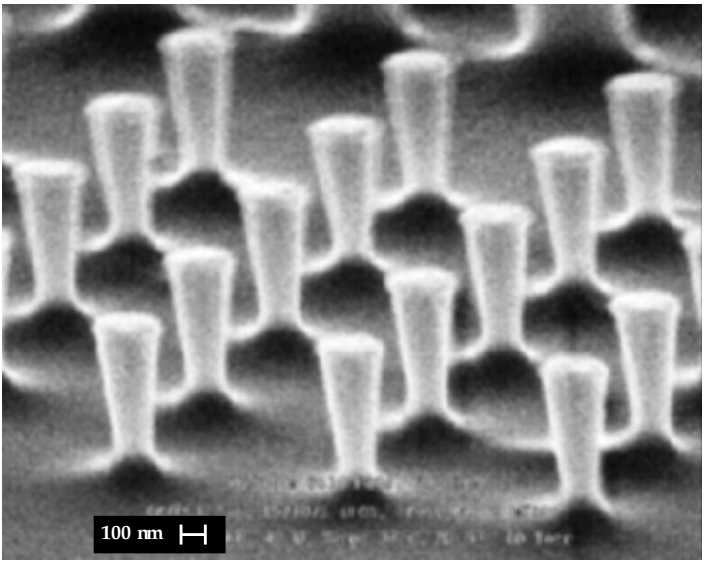

Fig. 11.2. Typical post structure of p–i–n junction that can be fabricated using the fabrication technology described in this section

rate of electrons with ions is small compared to the cyclotron frequency. The generated density of plasma is determined by the microwave power. Unlike conventional RIE, an ECR plasma does not generate a high bias voltage toward the sample chuck. An independent RF power source can be applied to the chuck, which controls the DC acceleration voltage for the ions to hit the sample. Since the RF power is not used to generate the plasma, the RF power can be decreased arbitrarily to decrease the acceleration voltage. In an ECR plasma, there are two independent parameters to control the plasma density (microwave power) and the DC bias to the sample (RF power).

In the fabrication process used for the mesoscopic p–i–n junctions, the following parameters were used: 15 sccm (unit of gas flow: standard cubic centimeters per minute) of Ar, 10 sccm of BCl_3, and 1 sccm of Cl_2 were supplied to the chamber, and the pressure was maintained at 2 mtorr. 400 W of microwave power was used, and 35 W of RF power was used to generate 45 V of DC voltage. The etch rate was about 1200 Å/min. Figure 11.2 shows typical post structures that can be fabricated using the electron-beam lithography and ECR-RIE technology.

11.2.6 Surface Passivation

One major problem with the idea of a small (sub-micron size) p–n junction light-emitting device made of GaAs is the high nonradiative surface recombination. It is well-known that there is a high density of surface states in GaAs very close to the middle of the energy gap (0.8 eV below the bottom of the conduction band) [155, 221]. Such states not only pin the Fermi energy at the middle of the gap at the semiconductor–air interface, but also act as nonradiative recombination centers for electron–hole pairs. The presence of surface states can be overcome in certain applications involving only one type of carrier (unipolar devices), but always puts a limitation on the performance of devices where both type of carriers are necessary (bipolar devices). For instance, in a metal–semiconductor field- effect transistor (MESFET), the metal gate deposited on the semiconductor surface forms a Schottky contact. The high density of electrons from the metal always saturates the surface states. Since there is no charging/discharging of the surface states on a short time scale, these states simply act as a surface charge density and do not affect the transistor performance at high frequencies [92]. At lower frequencies, charging and discharging of such states is known to produce the well-known $1/f$ noise. This is the reason why metal–insulator–semiconductor field effect transistors do not perform well with GaAs, since the surface states are not saturated. On the other hand, leakage current mediated by the surface states between the collector and the base is the limiting factor for the performance of heterojunction barrier transistors (HBTs) made of n–p–n (or p–n–p) GaAs/AlGaAs heterostructures [167, 201].

Nonradiative recombination of electron–hole pairs through surface states is the major obstacle in achieving high-performance light-emitting devices. The threshold of many edge-emitting and surface-emitting semiconductor lasers are limited by surface leakage current. Low threshold in transverse-junction- stripe (TJS) lasers is achieved by their unique design, where exposure of the p–n junction (and thus the active region) to surface is minimal [180]. Overcoming surface recombination has been the major theme in lowering the threshold of vertical-cavity surface-emitting lasers (VCSELs); the best performance so far has been achieved by confining the injection current to the center of the mesa structure by means of an oxidized AlAs layer [52, 61, 207]. Microlasers using whispering-gallery modes of a semiconductor disc also suffer from surface recombination [96]. The surface recombination is characterized by a surface recombination velocity, $S_{\rm v}$ (in cm/s), defined from the boundary condition (at $x = 0$) [222]

$$eD_{\rm a}\left.\frac{\partial n_{\rm a}}{\partial x}\right|_{\rm x=0} = eS_{\rm v}[n_{\rm a}(0) - n_{\rm a0}] \quad , \tag{11.1}$$

where $n_{\rm a}$ and $D_{\rm a}$ are the electron density and its corresponding diffusion constant inside the intrinsic active region, respectively. The surface recombination velocity is given by [222]

$$S_{\mathrm{v}} = \sigma_{\mathrm{n}} v_{\mathrm{th}} N_{\mathrm{st}} [n_{\mathrm{a}}(0) - n_{\mathrm{p0}}] \quad , \tag{11.2}$$

where σ_{n} is the capture cross-section of the electron by the surface trap state, v_{th} is the thermal velocity of the electrons and N_{st} is the number of surface trap states per unit area.

The surface states in GaAs are known to be formed by surface oxides or an inbalance in stoichiometry. The major direction in trying to eliminate the surface states has been to replace the oxygen atoms at the surface with other group-VI elements, such as sulfur or selenium, to improve both the electrical and optical performances of devices [112, 202, 214, 258]. Instead of the deep donor states for the oxygen, these elements form shallow donor states, and the recombination rate through these states can be significantly reduced. However, the binding energy of oxygen with the GaAs surface is much larger than that of sulfur, and the surface sulfides are eventually replaced by oxides in ambient conditions. Reduction of the surface recombination velocity after treatment in sulfide solution has been reported, but the improved performance only lasted for a short period of time [243]. One solution to this was to cover the sulfur-passivated surface with a silicon nitride film, which was stable for extended periods of time [96, 114]. Such passivation techniques report a reduction in surface recombination velocity of one to three orders of magnitude .

Using the well-known number of $S_{\mathrm{v}} = 10^6$ cm/s [114], the nonradiative recombination lifetime for an electron–hole pair through surface states is estimated to be ∼1 ps for sub-micron structures. Since the radiative recombination lifetime is on the order of $1 \sim 10$ ns, this means that there is no way that an electron–hole pair will generate a photon in such a sub-micron structure unless the surface recombination rate is significantly reduced.

We adopted the sulfur-passivation technique followed by silicon nitride encapsulation. First, the etched sub-micron post structures (Fig. 11.2) were dipped into an ammonium sulfide solution [$(NH_4)_2S$] heated to 60 °C and containing some excess sulfur, for 5 min. Then, the sample was rinsed in water and immediately put in a vacuum chamber, where a 500-Å-thick silicon nitride film was deposited using the plasma-enhanced chemical vapor deposition (PECVD) technique. PECVD is a conformal deposition technique, so the silicon nitride film covers the side of the post structure. The PECVD was performed in the same ECR system that was used to perform reactive ion etching, at an extraordinarily low-temperature of 100 °C. Such low temperature deposition is critical for GaAs and was made possible by the high-density plasma generated using the ECR technique.

An experiment was performed to demonstrate the effectiveness of the surface-passivation technique described above. Figure 11.3 shows the current–voltage characteristics of large area (100μm×100μm) p–i–n junctions prepared from identical wafers with different surface preparations, measured at 4 K. The dashed line shows the sample prepared by wet chemical etching in H_3PO_4/H_2O_2 solution. One can see that a significant amount of current

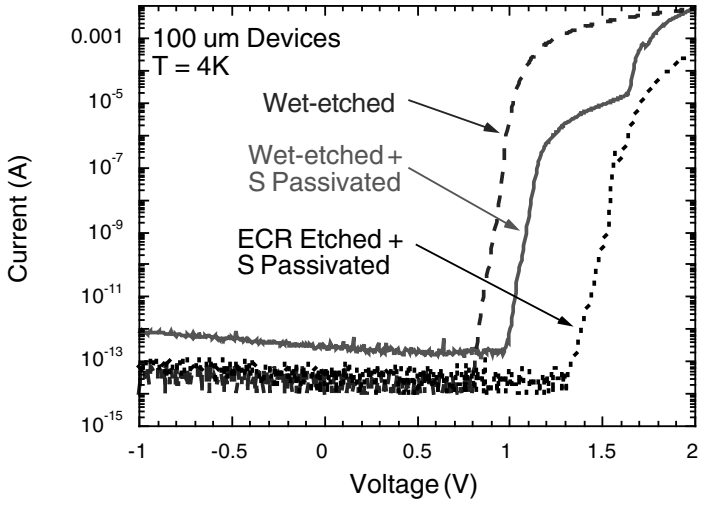

Fig. 11.3. Current–voltage characteristics of p–n junctions with different surface preparations. The *dashed*, *solid* and *dotted lines* correspond to wet-etched, wet-etched + sulfur-passivated, and ECR-etched + sulfur-passivated samples, respectively

starts to flow at a DC bias voltage of 0.8 V. Since the current should not flow in a p–n junction until the bias voltage reaches the built-in potential (∼1.5 V for GaAs at 4 K), the current below 1.5 V is attributed to the leakage current through the surface. The solid line shows the current–voltage characteristics when this wet-etched device is treated by the sulfur-passivation technique described above. The leakage current decreases by about two orders of magnitude at the voltages below the built-in potential, consistent with other reports. When the surface is prepared by ECR etching and sulfur passivation, the result shown by the dotted line is obtained. The leakage current in this device is significantly reduced until the bias voltage approaches 1.5 V. The surface leakage current at voltages below this value is suppressed by more than 10 orders of magnitude in these samples, compared to the conventional method of wet etching.

The significant reduction of surface leakage current demonstrated in this experiment, compared to reports using simple sulfur passivation treatment, deserves further investigation. One conclusion that can be drawn from this experiment is that it is more important to initially prepare a good surface rather than to try to recover a badly damaged surface in order to suppress the surface recombination. Wet etching of GaAs is achieved by oxidizing the GaAs material with H_2O_2 and removing the oxide with either an acid or a base. That process inherently leaves a high density of surface oxides, which result in a large surface recombination velocity. Dry etching using low-energy ECR plasma seems to leave an etched surface with low surface recombina-

tion rates. It was also found that annealing the ohmic contacts after the posts have been etched results in high surface recombination rates. This is because either the surface stoichiometry is altered at high temperatures (the GaAs crystal starts losing As atoms at the surface above 400 °C) or oxidation is enhanced at high temperatures. For the samples for which the ohmic contacts are annealed before the p–n junction surface is etched, and the etching is achieved by ECR-RIE, the surface recombination rates were always low. This conjecture is supported by experiments in micron-scale semiconductor microcavity lasers that were demonstrated recently. Assumption of an exceptionally low surface recombination velocity is necessary to explain the threshold values achieved in these devices [196, 199, 211]. The fact that these optically pumped devices did not need any annealing procedure, and that they were also etched using low-energy ECR-RIE or kinetically assisted ion beam etching (KAIBE), is consistent with the conjectures made here to describe our devices.

11.2.7 Planarization and Top-Contact Evaporation

Once the post structure is made and the surface of the p–n junction is properly passivated, the top of the post needs to be electrically contacted to the outside world. In order to achieve this, the structures were first planarized with some insulating material and large contact pads that can be wirebonded were put down. The insulator chosen is hard-baked photoresist, since it is easy to work with and readily available. Photoresist (Shipley 1805 by Microposit) was spun on at 4000 rpm (rotations per minute) to give a ∼5000 Å film. Since the thickness of the photoresist film is on the order of the lateral size of the posts, the thickness of the photoresist on top of the posts will be much thinner than 5000 Å when the photoresist is baked due to thermal flow. Since photoresist normally dissolves in acetone and another liftoff process is necessary to put down the contact pads, hard-baking is required to change the chemistry of the photoresist. An independent characterization experiment shows that when photoresist is baked at 150 ∼ 180°C for an extended period of time (>12 h in an oven, or >20 min on a hot plate) it becomes resistant to acetone. The electrical characteristics of such hard-baked photoresist is also tested as to whether it is insulating or not. If photoresist is baked at a much higher temperature (above 200°C), the film becomes metallic.

After covering the whole structure with hard-baked photoresist, the top of the posts need to be exposed for further electrical contact. This is achieved by a short plasma etch in O_2 to remove a thin layer of photoresist, and another short plasma etch in CF_4/O_2 to remove the silicon nitride film on top. This process exposes the top metal contact (which was used as the mask to etch the post structures), and another layer of metallization can be added to lead to large wirebonding pads. In this p–i–n junction device, the photons are extracted only through the top, since any photon injected to the substrate side is absorbed by the GaAs substrate. Therefore, the metal contact pad

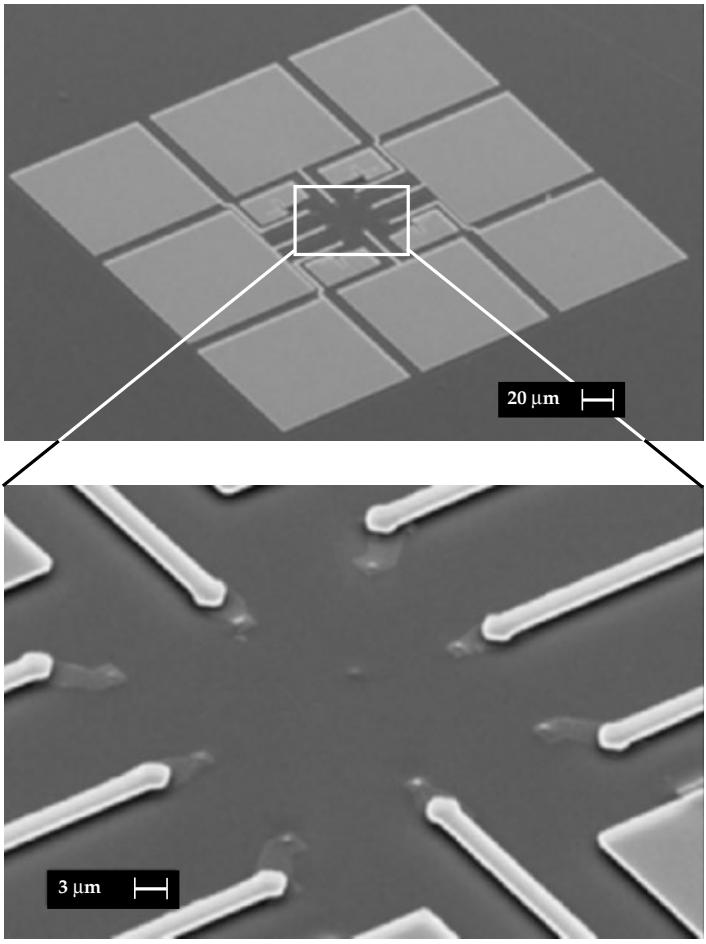

Fig. 11.4. An SEM micrograph of the final sub-micron p–i–n junction device. The large contact pads were 100 μm on one side and used for wirebonding. Eight contact pads are made in one set and are in contact with eight mesoscopic p–i–n junctions. The small pads on top of the posts allow photons to escape

should be thin enough to guarantee sufficient optical transparency. An aligned electron-beam-lithography step was performed to put down thin metal square pads (∼200 Å thick) that align to the top of the posts in one corner. Then, thick (∼3000 Å) gold contact pads were evaporated using optical lithography that was used for wirebonding. A scanning electron microscope (SEM) picture of this is shown in Fig. 11.4.

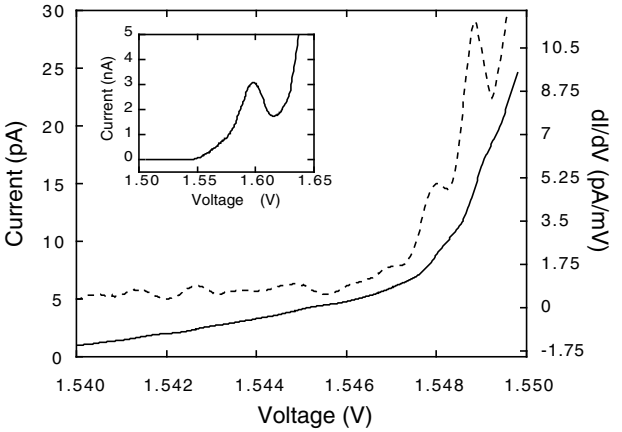

Fig. 11.5. Current–voltage characteristics (*solid line*) and $\mathrm{d}I/\mathrm{d}V$–voltage characteristics (*dashed line*) of a higher tunneling rate p–i–n junction at 50 mK. The I–V curve is measured near the lower tail of the resonant tunneling peak. The inset shows the I–V curve for a wider voltage range, in which the resonant tunneling peak can be seen

11.3 Observation of the Coulomb Staircase

In this section, evidence of the Coulomb blockade effect is presented in a mesoscopic p–i–n junction device fabricated using the procedure described above. The p–i–n junction structure studied in this section was slightly different from that described in Table 11.1. The junction used in this device was optimized for higher electron and hole tunneling rates and has lower (15 % Al concentration instead of 18 %) and thinner (160 Å thickness instead of 200 Å) barriers, and the p-side and n-side QWs are made of GaAs instead of AlGaAs. The fabricated devices were installed in the mixing chamber of a dilution refrigerator, at a base temperature of ∼50 mK. The junction was biased with DC and AC voltage sources, so that the DC current–voltage characteristics of the device could be measured with the option of an AC modulation signal superimposed on top.

Figure 11.5 shows the DC current–voltage characteristics of a device with a lithographic diameter of 600 nm at a very low bias voltage. The inset shows the measured I–V characteristics over a wider voltage range, in which the resonant tunneling peak can be seen. In this device at a low voltage, the hole tunneling rate curve has a very small slope, which can be derived from the I–V characteristics. As predicted by the model described in Chap. 10, small steps in the I–V curve can be observed, which show up as a periodic oscillation in the calculated $\mathrm{d}I/\mathrm{d}V$ curve. Each period corresponds to an addition of one electron in the central QW. From the period of approximately 1.1 ± 0.2 mV we derive a junction capacitance of 0.15 ± 0.03 fF.

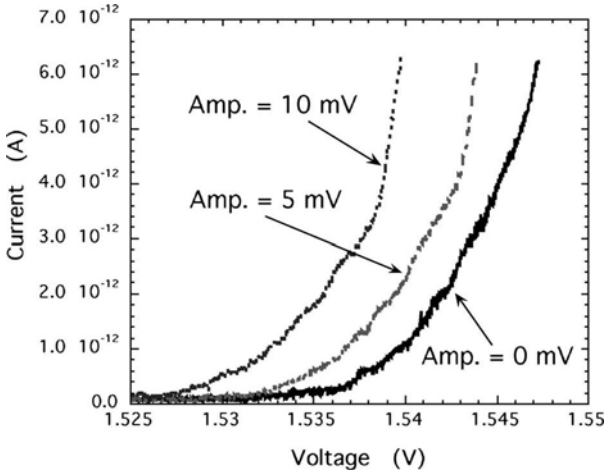

Fig. 11.6. Current-voltage characteristics of the same p-i-n junction as in Fig. 11.5 with an AC square-wave modulation. The frequency was fixed at 10 MHz, and the amplitude was varied to produce the different curves

The Coulomb staircase was also observed at this low bias voltage when a square-wave-modulation signal of frequency / = 10 MHz and amplitude ranging from 3 mV to 10 mV was applied. This is shown as the modulation of the DC current as a function of bias voltage in Fig. 11.6. This results from the fact that the modulation frequency is fast compared to the hole tunneling rate. In this situation the current is determined solely by the hole tunneling rate [(10.19) and (10.20)]. As in the case without modulation, the steps in the *I-V* characteristics reflect the charging of the QW by additional electrons. It was also confirmed experimentally that there was no frequency dependence of the Coulomb staircase steps in this limit.

The current-voltage measurements on this device show that the Coulomb blockade effect exists in a sub-micron p-i-n junction cooled to millikelvin temperatures. This is the first experimental demonstration of the Coulomb blockade effect between two types of carriers (electrons and holes) in a bipolar device.

11.4 Single-Photon Turnstile Device

11.4.1 Preliminary Characterization

Experimental evidence for single-photon turnstile operation was observed in a sample with the junction structure described in Table 11.1. Before searching for experimental evidence for single-photon turnstile device operation, a few measurements were performed to characterize the electrical and optical

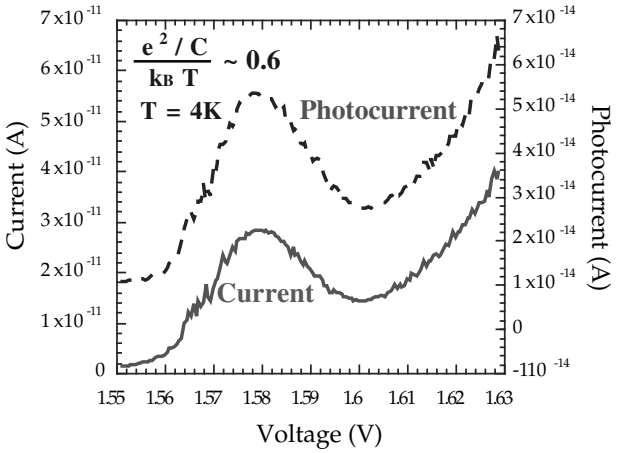

Fig. 11.7. Measured current (*solid line*) and photocurrent (*dashed line*) of a mesoscopic p–i–n junction with a diameter of 600 nm, fabricated by the procedure described in this chapter. The measurement was done at 4 K, and the photocurrent was measured with an avalanche photodiode

performance of the device. These preliminary measurements were done at 4 K.

First, the DC current–voltage characteristics was measured for a device with diameter of 600 nm. A well-defined resonant tunneling current peak with a very small background current was observed, as shown with a solid line in Fig. 11.7. This indicates that surface (leakage) current is suppressed well in spite of the very small size of the post created by the fabrication process described. The maximum value of the resonant tunneling current indicates a maximum electron tunneling rate of 2×10^8 s^{-1}, which is not very far from the estimated value of 7×10^7 s^{-1} calculated in Chap. 10.

The photons that are generated from this were was also measured, using a large area avalanche photodiode (APD; with an active area diameter of ∼5 mm). The APD was placed about 1 mm away from the surface of the mesoscopic junction, in a face-to-face configuration. The APD was also cooled down to 4 K, and the avalanche gain was about 60 at the bias point used. The dashed line in Fig. 11.7 shows the measured photocurrent through the APD normalized by the avalanche gain, to provide the primary photocurrent, I_p. One should note that the shape of I_p follows the shape of the device current, I_d, very closely. The relation between these two currents is given by

$$I_\mathrm{p} = \eta_\mathrm{i}\eta_\mathrm{e} I_\mathrm{d} \quad , \tag{11.3}$$

where η_i is the internal quantum efficiency for conversion of an electron–hole pair to a photon and η_e is the external coupling efficiency for the generated photon to reach the detector. From Fig. 11.7, one can conclude that the overall current-to-photocurrent conversion efficiency ($\eta_\mathrm{i}\eta_\mathrm{e}$) is about 2×10^{-3}.

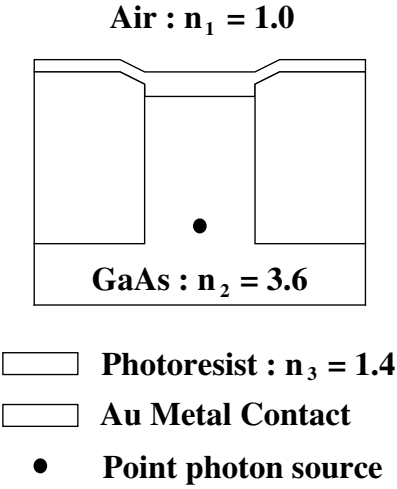

Fig. 11.8. Schematic of a model used to estimate the external coupling efficiency of a photon to escape the device and reach the photodetector

The external coupling efficiency (η_e) for a photon to escape the single photon turnstile device and reach the detector can be estimated from simple geometric considerations. Figure 11.8 shows the model used for this estimation. A model was considered where a point photon source is placed inside a post structure that is surrounded by photoresist and covered by a thin gold film. The thicknesses of the gold film used for the calculation were 500 Å on top of the posts and 300 Å on the photoresist surrounding the structure. The transmittivity of the photons (at $\lambda = 820$ nm) through the gold films was calculated [27] to be about 3.8 % on top of the posts and about 16 % on top of the photoresist. Such absorption loss due to the top metal contact is the main reason for reduced total external coupling efficiency. The other reason for low efficiency is the total internal reflection at the GaAs–air interface, due to a large index mismatch. For such a post structure, the total fraction of photons allowed to leave the GaAs crystal is limited to about 9 % due to total internal reflection. Accounting for these two loss mechanisms, the total external coupling efficiency is calculated to be less than $\sim 6 \times 10^{-3}$. From this calculation, a lower bound on the intrinsic electron-to-photon conversion efficiency (η_i) can be estimated to be $\geq$33 %.

11.4.2 Experimental Setup

Figure 11.9 shows the schematic of the experimental setup used in this experiment. The device was installed in a dilution refrigerator with a base temperature of $\sim$50 mK. Both a DC and an AC bias voltage were applied

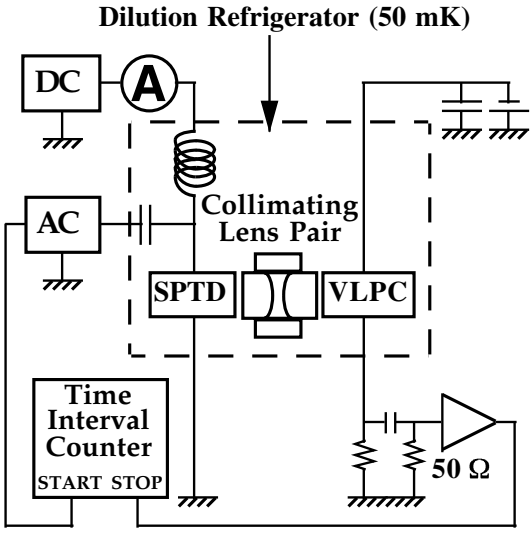

Fig. 11.9. Schematic of the experimental setup for the single-photon turnstile device

using a bias tee, also installed in the mixing chamber, less than 1 inch away from the sample. A DC current flowing through the device was measured as a function of DC bias voltage with a square-wave AC modulation voltage superimposed on top. It is very important to verify that the square-wave AC signal from the function generator is transferred to the sample without significant distortion. The bandwidth of the bias tee used was 100 kHz − 4.2 GHz. Figure 11.10a is the input pulse generated from our pulse generator, with about 45 ns on-pulse duration and 100 ns period. Figure 11.10b shows the pulse measured at room temperature after it passes through the bias network, including the coax cables inside and outside the dilution refrigerator, a 20 dB attenuator installed at the 1 K pot of the dilution refrigerator, and the bias tee placed at the mixing chamber. The input to the oscilloscope was placed about 1 inch away from the output of the bias tee in place of the p–n junction sample. The amplitude of the pulse is decreased by a factor of 10 due to the 20 dB attenuator that was installed to attenuate the room temperature thermal noise and provide good heat-sinking of the AC coax cable. A DC bias of 10 mV was also applied to show that applying DC bias does not affect the pulse shape in this bias tee.

The emitted photons from the mesoscopic p–i–n junction device were focused onto a single-photon counting detector based on a visible-light photon counter (VLPC) using a collimating lens pair. The characteristics of this detector are discussed in detail in Chap. 12. The detector's performance is optimized at a temperature of 6∼7 K. The VLPC was also installed in the mixing chamber of the dilution refrigerator, but the temperature was held at

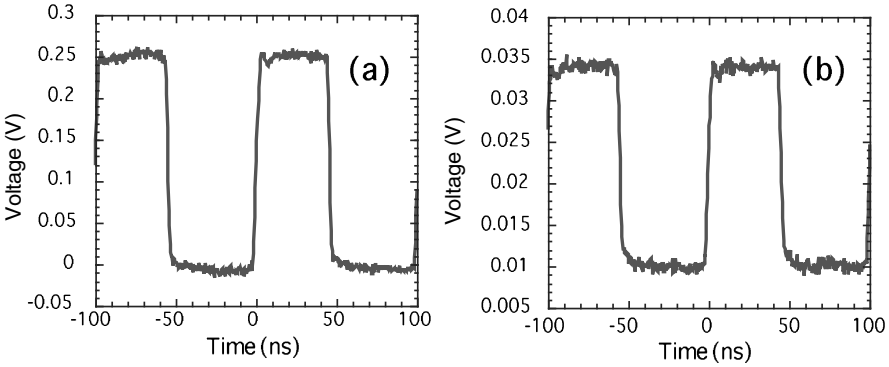

Fig. 11.10 a, b. Real time trace of an AC square wave at (**a**) the input and (**b**) the output of the bias network including coax cables inside and outside the dilution refrigerator, a 20 dB attenuator installed at 1 K pot of the dilution refrigerator, and the bias tee placed at the mixing chamber

6.5 K with active temperature control and good thermal isolation provided by eight thin superconducting wires that suspended the detector in vacuum. The detector generates an electrical pulse with a duration of 2 ns and a very well-defined height. This signal was used to trigger the "start" port of a time-interval counter. The "stop" port of the time-interval counter could be triggered either by the rising edge of the AC modulation signal to the p–i–n junction, to measure the correlation between the modulation signal and the output photon flux, or by another photon signal from the photon counter, to measure the intensity–intensity correlation function (or the second-order coherence function) of the output photon flux.

11.4.3 Electrical Characterization

Single-photon turnstile operation requires that the electrons and holes are injected into the active region in a regulated fashion. This means that only one electron and one hole are injected per modulation period and implies that there is a strong correlation between the DC current and the modulation frequency [154]. Figure 11.11 shows the measured current as a function of AC modulation frequency for a device with a diameter of 600 nm. A fixed AC amplitude of 72 mV is superposed on the three different DC voltages, $V =$ 1.545, 1.547 and 1.550 V. As the modulation frequency was increased, the DC current increased linearly as a function of the modulation frequency. The measured current was in close agreement with the relation $I = ef$, $I = 2ef$, and $I = 3ef$ (solid lines), when a frequency-independent background current was subtracted. This background current varies from device to device, and ranges from 0.5 pA to 6.5 pA. In Fig. 11.12, the slopes I/f from the current versus frequency curves were evaluated and plotted as a function of the DC bias voltage. It is seen that the slope increases discretely, creating plateaus

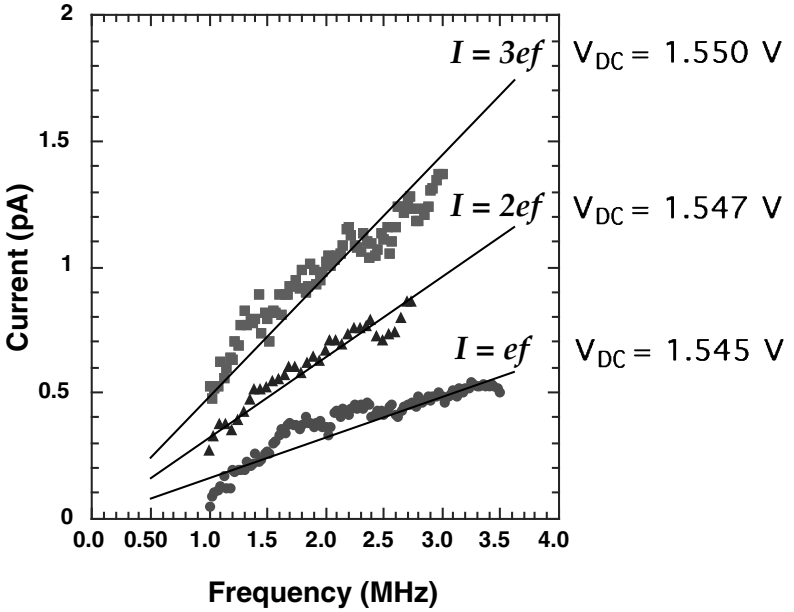

Fig. 11.11. The modulation frequency dependence of the DC current in the 600 nm turnstile device. The background DC (leakage) current, which is independent of the modulation frequency, has been subtracted. The measured current agrees with the relation $I = nef$ (*solid lines*), where e is the electron charge and $n = 1$, 2 and 3

at $I/f = ne$, where $e = 1.6 \times 10^{-19}$ C is the charge of an electron and $n = 1$, 2 and 3.

It is worthwhile mentioning that there is an optimum range ($\sim$72 $\pm$ 5 mV) for the amplitude ΔV of the AC modulation signal over which turnstile operation is sustained. This arises from the fact that the resonant tunneling in a realistic QW is never ideal, and the peak-to-trough ratio is always finite. If the modulation amplitude is too large, unwanted electron tunneling is allowed at high voltage ($V + \Delta V$) due to other leakage mechanisms, and the fraction of the regulated current decreases. The optimum modulation amplitude ($\sim$72 mV) is determined by the difference between the resonant tunneling conditions for electrons and holes, while the tolerance ($\pm$ 5 mV) is constrained by the decreasing ratio of the turnstile current to the (leakage) background current.

The locking of the current at multiples of the modulation frequency ($I = nef$) suggests that the charge transfer through the device is strongly correlated with the external modulation signal [50, 71, 130, 191]. At the first current plateau, $I = ef$, a single (the mth) electron and a single (the first) hole are injected into the central QW per modulation period, resulting in single- photon emission. At the second current plateau, $I = 2ef$, two [the

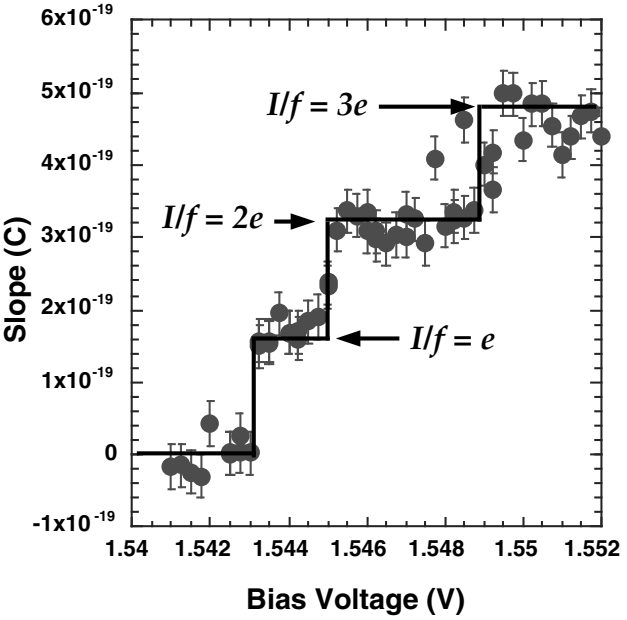

Fig. 11.12. The slopes I/f in the current–frequency curve versus DC bias voltage. The slopes I/f are quantized and forms plateaus at multiples of e, indicating operation at $I = ef$, $2ef$ and $3ef$

mth and $(m+1)$th] electrons and two (the first and second) holes are injected into the central QW per modulation period, resulting in two-photon emission. Similarly at the third current plateau, $I = 3ef$, three electrons and three holes are injected per modulation period, resulting in three-photon emission. This multiple charge operation becomes possible because of relatively broad inhomogeneous linewidths of the $n-$side and $p-$side QWs, as discussed in Chap. 10.

11.4.4 Optical Characterization

Having confirmed the regulated injection of electrons and holes into the active region, it is necessary to establish a correlation between the injected carriers and emitted photons, to confirm the turnstile operation for photons. To observe the time correlation between the modulation input and photon emission, we measured the time delay from the rising edge of the modulation input to the photon detection event at the first current plateau ($I = ef$) and the second current plateau ($I = 2ef$). The probability for a single electron–hole pair injected to the central QW of the turnstile device to be detected as a photon in the detector was about 1×10^{-4} due to a poor optical coupling efficiency between the two devices in the present setup. However, the quantum efficiency of detection does not affect the time correlation characteristics.

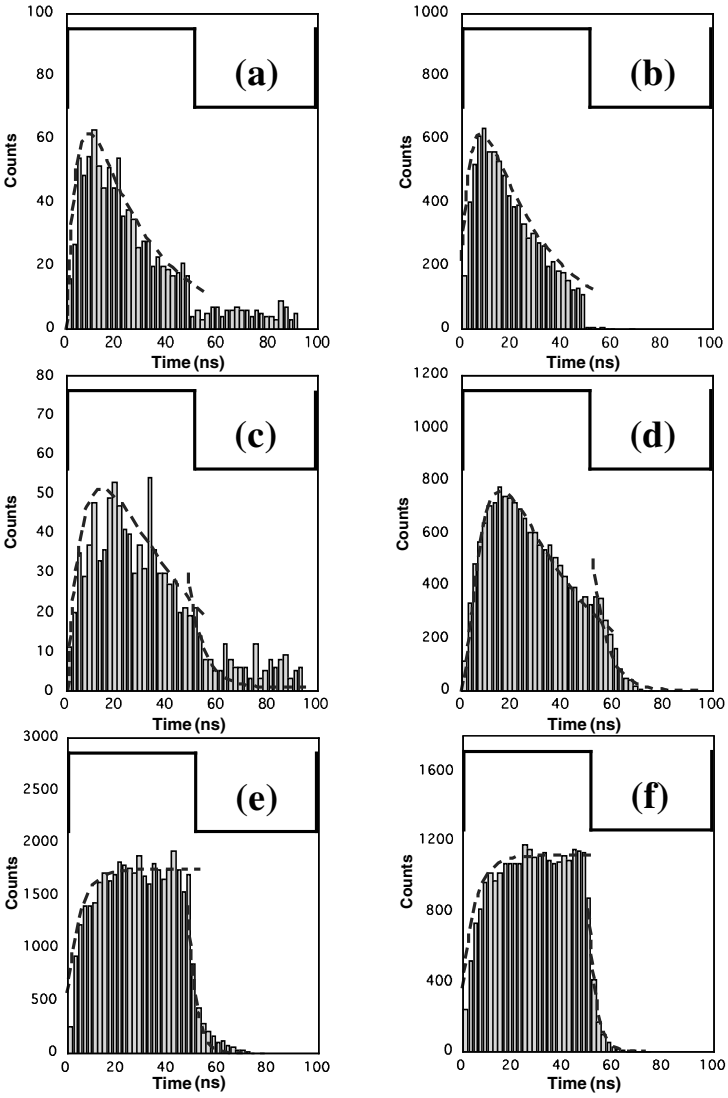

Fig. 11.13 a–f. Photon-emission characteristics of the turnstile device. (**a**) Measured histogram of the time delay between the rising edge of the modulation input and the photon-detection event at the first plateau ($I = ef$). (**b**) Monte-Carlo numerical simulation result for the first plateau. (**c**) Measured histogram of a time delay at the second plateau ($I = 2ef$). (**d**) Monte-Carlo numerical simulation result for the second plateau. (**e**) Measured histogram of a time delay for a larger-area device (diameter of 1.4 μm) at higher temperature (4 K), where the Coulomb blockade effect is absent. (**f**) Monte-Carlo numerical simulation result for this modulated classical light-emitting diode case

The histograms of the measured time delay with 10 MHz modulation frequency is shown in Fig. 11.13a (for $I = ef$) and Fig. 11.13c (for $I = 2ef$). The solid line on top indicates the external AC modulation voltage, and the finite photon counts during the off-pulse period are due to the dark counts of the VLPC. For the $I = ef$ plateau, a photon is emitted after the rising edge of the pulse since a single electron is injected at V_0 and a single hole is injected at $V_0 + \Delta V$. The photon-emission probability has a peak near the rising edge of the modulation input. The rapid increase of the photon-emission probability is associated with the hole tunneling time ($\tau_{\rm h} \simeq 4$ ns), and the slow decay of the photon-emission probability corresponds to the radiative recombination lifetime ($\tau_{\rm ph} \simeq 25$ ns). When only one electron and one hole are allowed to be injected and only one photon is emitted per modulation cycle, the photon-emission probability $P(t)$ for the time t after the turn-on of a higher voltage pulse is given by

$$P_1(t) = \frac{\exp(-t/\tau_{\rm ph}) - \exp(-t/\tau_{\rm h})}{\tau_{\rm ph} - \tau_{\rm h}} \quad , \tag{11.4}$$

where $\tau_{\rm ph}$ is the radiative recombination lifetime and $\tau_{\rm h}$ is the tunneling time of the hole. The dashed line is generated by this analytical formula using the parameters $\tau_{\rm h}$ and $\tau_{\rm ph}$ found from the experiment. Photon-emission probability in Fig. 11.13a decays to a non-zero value during the on-pulse due to photons generated by background current. The ratio of the counts contained in the peak to those contained in the non-zero background is ∼3:1, consistent with the ratio of the turnstile current to the background current in this device. Figure 11.13b shows the Monte-Carlo numerical simulation performed for this situation; it reproduces the experimental result well.

For the $I = 2ef$ plateau shown in Fig. 11.13c, the photon-emission probability distribution is broader, since two electrons and two holes are injected, and two photons are generated per modulation period. The photon-emission probability for the case where two electrons and two holes are allowed to tunnel can also be calculated analytically, as

$$\begin{aligned} P_2(t) = & \frac{2\tau_{\rm ph} - \tau_{\rm h2}}{2(\tau_{\rm ph} - \tau_{\rm h1})(\tau_{\rm ph} - \tau_{\rm h2})} \exp(-t/\tau_{\rm ph}) \\ & + \frac{2\tau_{\rm h1} - \tau_{\rm h2}}{2(\tau_{\rm h1} - \tau_{\rm h2})(\tau_{\rm h1} - \tau_{\rm ph})} \exp(-t/\tau_{\rm h1}) \\ & + \frac{\tau_{\rm h2}}{2(\tau_{\rm h2} - \tau_{\rm h1})(\tau_{\rm h2} - \tau_{\rm ph})} \exp(-t/\tau_{\rm h2}) \quad . \end{aligned} \tag{11.5}$$

Here, $\tau_{\rm h1}$ is the tunneling time for the first hole and $\tau_{\rm h2}$ is the (longer) tunneling time for the second hole. This theoretical curve is shown as a dashed line in Fig. 11.13c. The sharp cutoff of photon emission after the falling edge

of the modulation input is caused by the decay of the hole population due to simultaneous radiative recombination and reverse hole tunneling. The associated lifetime for this decay is $(\tau_{\rm h}^{-1} + \tau_{\rm ph}^{-1})^{-1}$. The experimental results as well as the analytical traces are well reproduced by the Monte-Carlo numerical simulation, as shown in Fig. 11.13d.

In order to compare this with photons from a classically modulated light-emitting diode, a control experiment is necessary where the Coulomb blockade effect is absent. Figure 11.13e shows the measured histogram of a time delay for a larger-area device (diameter of 1.4 μm) at higher temperature (4 K). In this case, an arbitrary number of holes are allowed to tunnel into the central QW during an on-pulse, and so the resulting photon-emission probability should increase monotonically to a steady-state value. This result is well-reproduced by the simulation, as shown in Fig. 11.13f. An analytical solution for the photon- emission probability can be found, and is given by

$$P_3(t) \;=\; \frac{C}{\tau_{\rm h}}\left[1 - \exp(-t/\tau_{\rm ph})\right] \quad , \tag{11.6}$$

where C is a normalization constant. This analytical solution was used to generate the dashed lines in Fig. 11.13e and f. The photon-emission probability increases monotonically through the duration of the on-pulse with the time constant $\tau_{\rm ph}$, as one would expect from a turn-on of a classical photon source. The radiative recombination time is shorter in this case ($\simeq$5 ns) due to a higher carrier density under the operating conditions.

The photon emission probabilities shown in Fig. 11.13a–d and (11.4) and (11.5) are in sharp contrast to an ordinary pulse-driven classical photon source where the Coulomb blockade is absent [shown in Fig. 11.13e and f and (11.6)]. The fact that the photon-emission probability decreases during the on-pulse duration is a unique signature that the number of holes injected during an on-pulse is restricted to either one or two due to the Coulomb blockade effect. This is a strong evidence for the Coulomb blockade effect and turnstile operation of our mesoscopic p–n junction.

11.5 Summary

In this chapter, the experimental effort towards generation of heralded single-photon states using the Coulomb blockade effect in a mesoscopic p–i–n junction was discussed. The process used to fabricate these small junctions was described, and the results of measurements performed at low temperatures were presented. The first experimental observation of a Coulomb staircase was reported in a system involving two different types of particles in a bipolar device.

Experimental evidence was also presented for the operation of a single-photon turnstile device. The electrical characterization demonstrates a frequency dependent current, which is a signature of time-correlated charge

transport due to the Coulomb blockade effect. The optical characterization, where we measured the photon-emission probability distribution with respect to the external modulation signal, implies that the number of carriers that are injected, and thus the number of photons that are emitted, are restricted due to the Coulomb blockade effect.

Such experimental evidence provides proof of the fact that the Coulomb blockade effect can be utilized to manipulate photon statistics and produce sub-Poissonian and antibunched photon states even down to a single-photon level. However, there are imperfections in the current devices that prohibit a direct demonstration of these nonclassical features. Improvements on the device performance is possible, with better confinement of both carriers and photons [90, 91].